

LSTM による文脈を考慮した論証マイニング手法の提案

木下 涼 宇都 雅輝 植野 真臣

電気通信大学大学院 情報理工学研究所

{kinoshita, uto, ueno}@ai.is.uec.ac.jp

1 まえがき

近年、自然言語処理において、文章中の論理構造を自動推定する論証マイニング (argument mining) と呼ばれる技術が注目されている [1, 4, 5, 6, 7, 8]. 論証マイニングでは、文章の論理構造は段落ごとに構築されているものとみなし、各段落について、以下の四つのサブタスクを解くことによって、論理構造を推定する。

要素抽出 文章中から、論理構造の構成要素となる文または節を論理要素として抽出する。

要素分類 各論理要素を「主張」あるいは「前提」に分類する。「主張」は、筆者の意見を明らかにする箇所であり、「前提」は、データや例を示すことにより他の要素の説得力を高めている箇所である。

関係分類 段落内の全ての論理要素間について因果関係の有無を推定する。ここでの因果関係とは、ある要素が別の要素の内容を支持しているような関係を示す。要素 i から要素 j に因果関係があると分類された時、要素 i から要素 j に有向エッジが引かれているとみなし、要素 i を原因要素、要素 j を結果要素と呼ぶ。

構造推定 要素分類と関係分類の結果を統合し、線形計画法などを用いて条件を満たすような構造を決定する。

論証マイニングは、情報検索や文章要約、自動採点、論文執筆支援など数多くの自然言語処理タスクに応用されており、重要性が高まる基礎技術と考えられる。

論証マイニングの四つのサブタスクのうち、要素抽出に関しては、ディープラーニング手法を用いることで高い精度が実現されている [1]. 一方、その他のサブタスクに関して最先端の手法 [8] では要素分類と関係分類に SVM を使い、構造推定を線形計画法で行う手法を提案しているが、以下の問題により大きな精度向上には至っていない。

- 1) 各要素が主張か前提かは前後の文脈に強く依存すると考えられるが、この手法では、要素分類に、前後の文脈を活用していない。
- 2) 関係分類において、要素分類の結果を考慮しておらず、独立したタスクとしてみなしている。

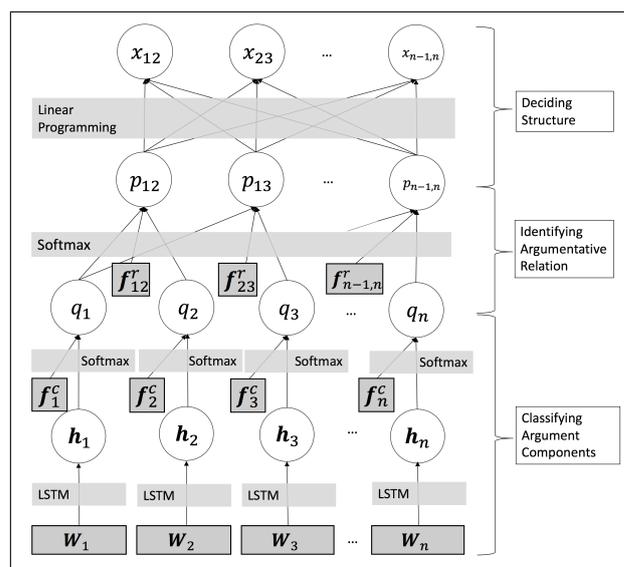


図1 提案手法の概念図

3) 要素分類と関係分類の分類結果を結合する際に、線形計画法の目的関数において各結果に対する重みが恣意的に決定されている。

本研究ではこれらの問題を解決するために、ディープラーニングの一種である long short-term memory (LSTM) を用いて文脈を考慮した論証マイニング手法を提案する。提案手法では、要素分類における特徴量として、対象要素の前文も用いて学習した LSTM の隠れ層の値を用いる。さらに、ソフトマックス関数によって得られる要素分類の結果を関係分類の特徴量として採用する。また、構造推定では、関係分類で推定された各要素間の因果の有無に関する確率を、線形計画問題の目的関数の重みとして採用する。

この手法には以下のような利点が存在する。

- ・ LSTM を用いて文脈情報を取り込むことにより、要素分類の精度が向上する。
- ・ 従来は独立に解かれていた要素分類と関係分類を、同時に解くことができる。
- ・ チューニングパラメータを排除することができ、構造の推定精度の改善が期待できる。

ベンチマークコーパス [8] を用いた評価実験によっ

て、提案手法では先行研究 [8] と比較して、要素分類における「主張」の分類精度が約 7%、関係分類における「因果関係有」の分類精度が約 8% 向上することが示された。また、提案手法の構造推定では、要素分類と関係分類の結果を適切に反映できることが示唆された。さらに、提案手法全体では、先行研究と比較して「因果関係有」の分類精度が約 7%、「因果関係無」の分類精度が約 2% 高いことが明らかになった。

2 提案手法

本研究では、LSTM を用いて文脈を考慮した論証マイニング手法を提案する。提案手法の概要を図 1 に示す。図中の \mathbf{W}_i は任意の段落の i 番目の論理要素とその前文の単語ベクトルであり、 \mathbf{h}_i は \mathbf{W}_i を入力とした時の LSTM の最後の隠れ層を表す。 \mathbf{f}_i^c は i 番目の要素に対する要素分類の特徴量ベクトルを示しており、 \mathbf{f}_{ij}^r は要素 i を原因要素から要素 j を結果要素とした時の特徴量である。要素分類では、 \mathbf{h}_i と \mathbf{f}_i^c を用いて、要素 i が「主張」である確率 q_i を求める。次に、関係分類では、 q_i , q_j , \mathbf{f}_{ij}^r を用いて、要素 i から要素 j に因果関係が存在する確率 p_{ij} を求める。最後に、それらの確率を目的関数とする線形計画法を用いて、木構造の制約のもと、要素 i から要素 j に因果関係が存在するとき 1 そうでなければ 0 をとる $x_{ij} \in \{0, 1\}$ の値を決定する。

提案手法には以下の特徴がある。

- 1) 要素分類において、LSTM の入力を前後の文章として得られた隠れ層の値を既存の特徴量と組み合わせて分類する。
- 2) 関係分類において、既存の特徴量に加えてソフトマックス関数によって得られる要素分類の結果を用いる。
- 3) 構造推定では関係分類で推定された各要素間の因果の有無に関する確率を線形計画問題の目的関数の重みとして用いることで、チューニングパラメータのない線形計画問題として定式化する。

以降では、各項目について詳細を説明する。

2.1 要素分類結果の改善

本研究では、文脈を考慮した要素分類を行うために、要素分類の特徴量として、文脈を考慮できるディープラーニング手法として知られる LSTM の隠れ層を用いる。自然言語処理分野においては、ディープラーニングの隠れ層を分類タスクに用いる手法が広く活用されており [2]、分類精度の改善に有効であることが示されてきた。LSTM は、時系列データに対するディープラーニング手法として近年自然言語処理分野を中心に広く活用がなされている。

LSTM には、時点 t における入力 i_t と時点 $t-1$ の隠れ層 h_{t-1} を用いて計算される 3 つのゲートベクトルが導入されている。これらのゲートベクトルの導入により長期依存を扱うことが可能になっている。LSTM の隠れ層は、長期の依存関係を反映したものであり、文脈の情報が保持されているとみなすことができる。

本手法において、LSTM の入力 i_t は論理要素及び文脈（前後の文章）における t 番目の単語である。そのため、最大の入力長は論理要素と文脈を含めた単語数となる。またパラメータの最適化アルゴリズムには adaptive moment estimation [3] を用いた。

LSTM の隠れ層を活用した分類問題では、Average Pooling や最後の隠れ層が用いられている [2]。そのため、LSTM の最後の隠れ層ベクトルを特徴量として用いる手法と全ての隠れ層の平均を取った Average Pooling を特徴量として用いる手法の二つを提案し、比較する。LSTM を用いた特徴量以外の特徴量は、Stab and Gurevych [8] と同じものを使用する。

2.2 関係分類結果の改善

Stab and Gurevych [8] の手法では、要素分類と関係分類を別々に解いているが、これらは互いに強く依存した問題と解釈できる。そこで、要素分類の結果を関係分類の特徴量として用いることで、要素分類の結果を関係分類に直接的に反映する手法を提案する。

提案手法では関係分類に「主張特徴」を新たに用いる。これは、要素分類においてソフトマックス関数を用いて求められる各要素が「主張」である値、あるいは「主張」であると判別されたかどうかの 2 値のどちらかの値である。

2.3 構造推定の改善

Stab and Gurevych [8] は、要素分類と関係分類の結果を線形計画法を用いて統合し、木構造を推定している。その際に用いられるチューニングパラメータは恣意的に決定されており、結果に大きな影響を及ぼしている。本研究では式 (6) のように関係分類結果を重みとする目的関数を設定し、線形計画法を用いて最終的な木構造を決定する。確率値を重みとすることで、チューニングパラメータを用いずに最適化を行うことができる。制約は既存研究 [8] と同様である。

$$\arg \max_x \sum_{i=1}^n \sum_{j=1}^n p_{ij} x_{ij} \quad (1)$$

ここで、 $x_{ij} \in \{0, 1\}$ は要素 i から要素 j に因果関係があるとき 1、そうでなければ 0 をとる変数であり、線形計画法によって値を決定する。 $p(r_{ij} = 1)$ は要素 $i-j$ 間に因果関係が存在する確率であり、ソフトマッ

クス関数によって求められる。この線形計画問題は、全ての要素間に確率付きのエッジが引かれたものを入力として、各エッジの確率の和が最大となる木構造を発見する問題と解釈できる。ここでは、恣意的なパラメータは存在せず、関係分類の結果を適切に反映できていると考えられる。

3 評価実験

本章では、提案手法によりそれぞれのサブタスクの結果がどのように変化するか評価する。評価実験においては、Argument Annotated Corpus[8]を用い、10交差検証を行って精度 (precision) と再現率 (recall) の調和平均である F 値を求めた。それぞれの比較に関しては比較対象である Stab and Gurevych[8] と統一し SVM を用いて分類している。

3.1 要素分類の評価実験

まずはじめに、要素分類に LSTM による特徴量を用いる効果を評価する。Stab2017 を再現したものと LSTM による特徴量を用いた提案手法に関して、要素分類の結果を比較した結果を表 1 に示す。LSTM への入力、対象となる要素とその前文 1 文 (Pf1)、その前文 2 文 (Pf2)、対象となる要素とその次文 1 文 (Pb1)、その次文 2 文 (Pb2) の 4 パターンを用いた。次文の文脈を用いる際は、より文脈の情報を考慮できるように単語列の順序を逆転した上で入力とした。隠れ層の次元数はそれぞれ、50・100・200 次元で比較している。

表 1 より、LSTM によって得られた隠れ層ベクトルを特徴量に加えることで「主張」及び「前提」のどちらに対しても既存の精度を上回ることがわかる。今回の実験では、Last_Pf1 の次元数 200 が最も高精度となった。しかし、2 文以上の文脈を含めた入力では、精度は大きく低下する傾向にあることがわかった。これは先行研究で報告されているように [9]、LSTM を用いても 3 文以上の入力に対しては学習が難しくなることが原因と考えられる。また、今回の結果からは、次文の文脈を考慮することによる結果はみられなかった。よって、適切な条件のもと LSTM を用いることで文脈の情報を十分に活用できていることがわかる。

3.2 関係分類の評価実験

次に、関係分類に要素分類の結果を用いる効果を評価する。ここでは、3.1 節の要素分類の結果をそのまま 2 値変数として特徴量に加える場合 (binary_LSTM) とソフトマックス関数による各要素が「主張」である値を出力し、関係分類の特徴量に加える場合 (softmax_LSTM) の関係分類の精度を求めた。また、関係分類を独立して行ったもの (Stab2017)、Stab2017 と

同じ特徴量を用いて要素分類を行った結果を用いて同様の実験を行ったもの (binary_Stab,softmax_Stab) とも比較している。これらの結果を表 2 に示す。

表 2 より、Stab2017 と比較して、softmax_Stab では精度が大きく向上していることがわかる。さらに、もっとも精度が高いものが softmax_LSTM であることから、3.1 節での改善が適切に反映されていることがわかる。

3.3 構造推定の評価実験

本節では、提案する構造推定が先行研究 [8] よりも適切な重みに基づく構造推定であることを比較実験によって示す。ここでは Stab2017 と全く同じ特徴量を用いた提案手法を比較した。また、要素分類に真の値を使った Stab2017(True) と提案 (True) についても同様に精度を求めた。構造推定について、各要素間の因果関係の有無の F 値を算出した結果を表 3 に示す。

表 3 から、既存の特徴量を用いた場合、提案手法よりも Stab2017 の精度が高いことがわかる。この原因としては Stab2017 が要素分類の影響を受けにくい定式化であるため、現状問題となっている要素分類の精度の影響が小さいことが考えられる。その証拠として、要素分類の値にコーパスの値を用いた場合、提案手法は先行研究の構造推定よりも「因果関係有」の分類精度が大きく上回ることがあげられる。従って、提案手法はより要素分類の結果を適切に用いることができる手法であることがわかる。

3.4 提案手法全体の評価実験

最後に、これまでの 3 つの精度改善を統合し、Stab2017 と比較する。Stab2017 は先行研究をそのまま再現したものであり、提案手法ではこれまでの結果を踏まえて以下のような値を用いている。1) 要素分類の LSTM の入力には前文 1 文を用い、中間層は 200 次元で最後の層のみを用いる。2) 要素分類ではソフトマックス関数を用いて各要素が「主張」である値を求め、それを関係分類の特徴量として用いる。3) 要素分類の結果を適切に反映するため構造推定には提案手法を用いる。これらの比較結果を表 4 に示す。

表 4 より、既存手法では因果関係ありの分類精度が大幅に精度が向上していることがわかる。また因果関係なしの分類に関しても精度が向上している。以上より、提案手法によって論理構造をより高精度に推定できることが明らかとなった。

4 むすび

本論文では、より精度の高い論理構造の推定方法である論証マイニング手法を提案した。具体的には、(1)

表 1 LSTM を用いた比較実験結果

	入力	次元	F-Claim	F-Premise
Stab2017			0.627	0.832
提案 : Average	Average_Pf1	50	0.663	0.853
		100	0.671	0.858
		200	0.679	0.860
	Average_Pf2	50	0.610	0.823
		100	0.611	0.824
		200	0.605	0.825
	Average_Pb1	50	0.613	0.835
		100	0.612	0.836
		200	0.599	0.831
	Average_Pb2	50	0.609	0.830
		100	0.606	0.829
		200	0.606	0.827

	入力	次元	F-Claim	F-Premise
提案 : Last	Last_Pf1	50	0.608	0.832
		100	0.642	0.849
		200	0.698	0.875
	Last_Pf2	50	0.618	0.823
		100	0.601	0.831
		200	0.601	0.834
	Last_Pb1	50	0.615	0.834
		100	0.612	0.836
		200	0.598	0.834
	Last_Pb2	50	0.613	0.835
		200	0.601	0.832
		200	0.596	0.833

表 2 関係分類の評価実験結果

	F-Relation	F-No Relation
Stab2017	0.428	0.731
binary_Stab	0.398	0.702
softmax_Stab	0.460	0.794
binary_LSTM	0.493	0.819
softmax_LSTM	0.507	0.828

表 3 構造推定の評価実験結果

	F-Relation	F- NoRelation	F-All
Stab2017	0.466	0.881	0.634
提案手法	0.451	0.879	0.627
Stab2017(True)	0.692	0.918	0.805
提案 (True)	0.770	0.954	0.865

表 4 提案手法全体の評価

	F_Relation	F_NoRelation	F_All
Stab2017	0.466	0.881	0.634
提案手法	0.539	0.897	0.718

要素分類の精度を向上させるために文脈を考慮できる LSTM を用いた特徴量を提案し, (2) 要素分類の結果を関係分類における特徴量として使用し, (3) 線形計画法の目的関数に関係分類の結果を用いることでチューニングパラメータを用いない形で定式化した。

コーパスを用いた評価実験により, それぞれの提案が効果的であること, 及び全てを統合した結果が先行研究の精度を大きく上回ることが明らかとなった。ただし, それでもまだ因果関係ありについての分類精度はまだ十分に高いとはいえず, 応用のためにはさらなる精度の向上が必要である。そのために, さらなるディープラーニング手法の活用やウェブ上の豊富な資源を活用し精度を向上させていく。

参考文献

[1] Yamen Ajjour, Wei-Fan Chen, Johannes Kiesel, Henning Wachsmuth, and Benno Stein. Unit segmentation of argumentative texts. In *Proceedings*

of the 4th Workshop on Argument Mining, pp. 118–128, Copenhagen, Denmark, 2017.

- [2] Yoon Kim. Convolutional neural networks for sentence classification, 2014.
- [3] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.
- [4] Huy V. Nguyen and Diane J. Litman. Context-aware argumentative relation mining. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1127–1137, Berlin, Germany, 2016. Association for Computational Linguistics.
- [5] Andreas Peldszus. Towards segment-based recognition of argumentation structure in short texts. In *Proceedings of the First Workshop on Argumentation Mining*, pp. 88–97, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
- [6] Isaac Persing and Vincent Ng. End-to-end argumentation mining in student essays. In *Proceedings of Human Language Technologies: The 2016 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 1384–1394, 2016.
- [7] Christian Stab and Iryna Gurevych. Identifying argumentative discourse structures in persuasive essays. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pp. 46–56. Association for Computational Linguistics, October 2014.
- [8] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, Vol. 43, No. 3, pp. 619–659, 2017.
- [9] Martin Sundermeyer, Ralf Schl ijter, and Hermann Ney. Lstm neural networks for language modeling, In INTERSPEECH, 2010.